

What is claimed is:

1. A computer implemented process to identify at least one pattern and its distribution in a set of data for the purpose of interpreting the data, the process comprising:

(a) representing a set of data by an original data matrix D residing in a storage device, and;

(b) decomposing the set of data into a set of patterns represented by a matrix F and their distribution represented by a matrix A, wherein the matrix F represents the set of patterns needed to describe the data and the matrix A represents the distribution of the set of patterns within the data matrix D, the decomposing comprising performing a Bayesian-based Monte Carlo calculation using at least the data matrix D to determine the matrices A and F, wherein the matrices A and F reconstruct the data matrix D and are more amenable to analysis than the data matrix D.

2. A process according to claim 1 further comprising:

(c) determining by Monte Carlo sampling the uncertainties of all values in the elements of matrix F and matrix A.

3. A process according to claim 1 wherein the decomposing is performed such that the combined number of the elements in the matrices A and F are significantly smaller than the number of elements of the original data matrix, and the uncertainties in the matrices A and F combine to yield the correct uncertainty in matrix D, the significantly smaller number of elements making the matrices A and F more amenable to analysis than the data matrix D.

4. A process according to claim 1 further comprising:

(c) using a statistical process to determine the number of independent patterns required to reconstruct the original data matrix D within a noise level from the subordinate matrices A and F.

5. A process according to claim 4 wherein the independent patterns are spectral shapes.

6. A process according to claim 5 wherein the statistical process is principal component analysis, the process further comprising:

(c) using the principal component analysis to correct for any instrumental frequency or phase shifts which appear in spectra of the original data matrix D.

7. A process according to claim 5 wherein rows of the original data matrix D are chemical shift imaging spectra associated with specific locations in a living organism, rows of matrix F are individual nuclear magnetic resonance (NMR) spectra associated with different tissue types, and rows of matrix A are amounts of each tissue type at each specific location within the living organism.

8. A process according to claim 5 wherein rows of the original data matrix D are NMR spectra associated with specific time points during an observation of a living organism, rows of matrix F are individual NMR spectra associated with different chemical species, and rows of matrix A are amounts of each chemical species at each time point.

9. A process according to claim 1 wherein rows of the original data matrix D are NMR recovery curves associated with specific locations within a living organism, rows of matrix F are individual NMR recovery curves associated with different tissue types, and rows of matrix A are amounts of each tissue type at each specific location within the living organism.

10. A process according to claim 1 wherein rows of the original data matrix D are levels of expression of individual messenger RNA (mRNA) species at specific times, rows of matrix F are patterns of physiologically related mRNA expression, and rows of matrix A are amounts of each expression pattern at each specific point in time.

11. A process according to claim 10 further comprising:

(c) measuring the mRNA levels by adding a detectable label to DNA derived from the mRNA; and

(d) quantitating the amount of label associated with the DNA as a measure of the mRNA levels.

12. A process according to claim 11 wherein the label is selected from the group consisting of a radioactive label and a non-radioactive label.

13. A process according to claim 10 wherein expression of the mRNA is measured by synthesizing a DNA molecule which is complementary to the mRNA and detecting the amount of DNA synthesized.

14. A process according to claim 13 wherein the DNA molecule is synthesized in a reverse transcriptase reaction.

15. A process according to claim 13 wherein the amount of DNA synthesized is measured by:

(c) adding a detectable label to the DNA; and

(d) quantitating the amount of label associated with the DNA as a measure of the amount of DNA synthesized.

16. A process according to claim 15 wherein the label is selected from the group consisting of a radioactive label and a non-radioactive label.

17. A process according to claim 10 wherein expression of the mRNA is measured by amplifying the mRNA to DNA and detecting the amount of DNA so amplified.

18. A process according to claim 17 wherein the amplifying is conducted in a polymerase chain reaction.

19. A process according to claim 10 wherein the mRNA levels are measured using an array.

20. A process according to claim 19 wherein the array is a high density gene chip array.

21. The process according to claim 19 wherein the array is a low density array.

22. The process according to claim 21 wherein the low density array is a filter or a plate array.

23. A process according to claim 1 wherein rows of the original data matrix D are levels of expression of individual messenger RNA (mRNA) species at specific locations within a living organism, rows of matrix F are patterns of

physiologically related mRNA expression, and rows of matrix A are amounts of each expression pattern at each specific location in the organism.

24. A process according to claim 23 further comprising:

5 (c) measuring the mRNA levels by adding a detectable label to DNA derived from the mRNA; and

(d) quantitating the amount of label associated with the DNA as a measure of the mRNA levels.

25. A process according to claim 24 wherein the label is selected from the group consisting of a radioactive label and a non-radioactive label.

10 26. A process according to claim 23 wherein expression of the mRNA is measured by synthesizing a DNA molecule which is complementary to the mRNA and detecting the amount of DNA synthesized.

27. A process according to claim 26 wherein the DNA molecule is synthesized in a reverse transcriptase reaction.

15 28. A process according to claim 26 wherein the amount of DNA synthesized is measured by

(c) adding a detectable label to the DNA; and

(d) quantitating the amount of label associated with the DNA as a measure of the amount of DNA synthesized.

20 29. A process according to claim 28 wherein the label is selected from the group consisting of a radioactive label and a non-radioactive label.

30. A process according to claim 23 wherein expression of the mRNA is measured by amplifying the mRNA to DNA and detecting the amount of DNA so amplified.

25 31. A process according to claim 30 wherein the amplifying is conducted in a polymerase chain reaction.

32. A process according to claim 23 wherein the expression of mRNA is measured using an array.

30 33. A process according to claim 32 wherein the array is a high density gene chip array.

34. The process according to claim 32 wherein the array is a low density array.

35. The process according to claim 34 wherein the low density array is a filter or a plate array.

5 36. A process according to claim 1 wherein rows of the original data matrix D are amounts of individual DNA species in specific individuals, rows of matrix F are patterns of physiologically related DNA species, and rows of matrix A are amounts of each DNA pattern in each individual.

10 37. A process according to claim 36 wherein the amount of DNA is measured by hybridizing to the DNA a complementary DNA having a detectable label attached thereto and measuring the amount of label so hybridized as a measure of the amount of DNA.

38. A process according to claim 37 wherein the label is selected from the group consisting of a radioactive and a non-radioactive label.

15 39. A process according to claim 36 wherein the amount of individual DNA is measured by synthesizing a DNA copy of the DNA to generate a synthesized DNA, wherein the synthesized DNA has a detectable label attached thereto and measuring the amount of label in the synthesized DNA as a measure of the amount of DNA.

20 40. The process according to claim 39 wherein the label is selected from the group consisting of a radioactive label and a non-radioactive label.

41. A process according to claim 36 further comprising:

(c) measuring the amount of DNA (non-amplified DNA) by amplifying the DNA (amplified DNA) in the presence of a detectable label; and

25 (d) measuring the amount of label associated with the amplified DNA as a measure of the amount of non-amplified DNA.

42. A process according to claim 41 wherein the detectable label is selected from the group consisting of a radioactive label and a non-radioactive label.

30 43. A process according to claim 41 wherein the amplifying is conducted by a polymerase chain reaction.

44. A process according to claim 36 wherein the amount of individual DNA is measured on an array.

45. A process according to claim 44 wherein the array is a high density gene chip array.

5 46. The process according to claim 44 wherein the array is a low density array.

47. The process according to claim 46 wherein the low density array is a filter or a plate array.

10 48. A process according to claim 1 wherein rows of the original data matrix D are amounts of individual DNA species at specific locations in a living organism, rows of matrix F are patterns of physiologically related DNA species, and rows of matrix A are amounts of each DNA pattern at each specific location in the organism.

15 49. A process according to claim 48 wherein the amount of DNA is measured by hybridizing to the DNA a complementary DNA having a detectable label attached thereto and measuring the amount of label so hybridized as a measure of the amount of DNA.

50. A process according to claim 49 wherein the label is selected from the group consisting of a radioactive and a non-radioactive label.

20 51. A process according to claim 48 wherein the amount of individual DNA is measured by synthesizing a DNA copy of the DNA to generate a synthesized DNA, wherein the synthesized DNA has a detectable label attached thereto and measuring the amount of label in the synthesized DNA as a measure of the amount of DNA.

25 52. The process according to claim 51 wherein the label is selected from the group consisting of a radioactive label and a non-radioactive label.

53. A process according to claim 48 further comprising:

(c) measuring the amount of DNA (non-amplified DNA) by amplifying the DNA (amplified DNA) in the presence of a detectable label; and

(d) measuring the amount of label associated with the amplified DNA as a measure of the amount of non-amplified DNA.

54. A process according to claim 53 wherein the detectable label is selected from the group consisting of a radioactive label and a non-radioactive label.

55. A process according to claim 53 wherein the amplifying is conducted by a polymerase chain reaction.

56. A process according to claim 48, wherein the amount of individual DNA is measured on an array.

57. A process according to claim 56 wherein the array is a high density gene chip array.

58. The process according to claim 56 wherein the array is a low density array.

59. The process according to claim 58 wherein the low density array is a filter or a plate array.

60. A process according to claim 1 wherein rows of the original data matrix D are amounts of individual DNA species at different times in a living organism, rows of matrix F are patterns of physiologically related DNA species, and rows of matrix A are amounts of each expression pattern at each specific point in time.

61. A process according to claim 60 wherein the amount of DNA is measured by hybridizing to the DNA a complementary DNA having a detectable label attached thereto and measuring the amount of label so hybridized as a measure of the amount of DNA.

62. A process according to claim 61 wherein the label is selected from the group consisting of a radioactive and a non-radioactive label.

63. A process according to claim 60 wherein the amount of individual DNA is measured by synthesizing a DNA copy of the DNA to generate a synthesized DNA, wherein the synthesized DNA has a detectable label attached thereto and measuring the amount of label in the synthesized DNA as a measure of the amount of DNA.

64. The process according to claim 63 wherein the label is selected from the group consisting of a radioactive label and a non-radioactive label.

65. A process according to claim 64 further comprising:

5 (c) measuring the amount of DNA (non-amplified DNA) by amplifying the DNA (amplified DNA) in the presence of a detectable label; and

(d) measuring the amount of label associated with the amplified DNA as a measure of the amount of non-amplified DNA.

66. A process according to claim 65 wherein the detectable label is selected from the group consisting of a radioactive label and a non-radioactive label.

10 67. A process according to claim 65 wherein the amplifying is conducted by a polymerase chain reaction.

68. A process according to claim 60 wherein the amount of individual DNA is measured on an array.

15 69. A process according to claim 68 wherein the array is a high density gene chip array.

70. The process according to claim 68 wherein the array is a low density array.

71. The process according to claim 70 wherein the low density array is a filter or a plate array.

20 72. The process according to claim 1 wherein rows of the original data matrix D are measurements of individual samples comprising mixtures of chemical compounds, rows of matrix F are the measurements associated with a single chemical compound, and rows of matrix A are amounts of each chemical compound in each of the individual samples.

25 73. The process according to claim 72 wherein the rows of the data matrix D are gas chromatography/mass spectra (GCMS) measurements, and the rows of matrix F are the GCMS spectra for the individual chemical compounds.

30 74. The process according to claim 72 wherein the rows of the data matrix D are infrared spectroscopy measurements, and the rows of matrix F are the infrared spectra for the individual chemical compounds.

75. The process according to claim 72 wherein the rows of the data matrix D are optical absorption spectroscopy measurements, and the rows of matrix F are the optical absorption spectra for the individual chemical compounds.

5 76. The process according to claim 72 wherein the rows of the data matrix D are fluorescence spectroscopy measurements, and the rows of matrix F are the fluorescence spectra for the individual chemical compounds.

77. The process according to claim 72 wherein the rows of the data matrix D are high pressure liquid chromatography/standard detection measurements, and the rows of matrix F are the spectra for the individual chemical compounds,
10 wherein the spectra are selected from the group consisting of GCMS spectra, infrared spectra, optical absorption spectra and fluorescence spectra.

78. The process according to claim 1 wherein at least one pattern is an amount of goods or services.

79. The process according to claim 1, wherein the rows of the data
15 matrix D are amounts of goods and services at various times, the rows of matrix F are the patterns of goods and services, and the rows of matrix A are a measure of how the amounts of goods and services are distributed over time.

80. The process according to claim 1, wherein the rows of the data
20 matrix D are amounts of goods and services at various locations, the rows of matrix F are the patterns of goods and services, and the rows of matrix A are a measure of how the amounts of goods and services are distributed over various locations.

81. The process according to claim 1 wherein at least one pattern is a monetary value.

82. The process according to claim 1 wherein the pattern distribution is
25 across entities.

83. The process according to claim 1 wherein the pattern distribution is across a space or a location.

84. The process according to claim 1 wherein the pattern distribution is across time.

85. The process according to claim 1 wherein representing a set of data by an original data matrix D involves counting a number of occurrences of events within the set of data and encoding the number of occurrences into the original data matrix D.

5 86. The process according to claim 1 wherein the original data matrix D is a set of spatially dependent functions, matrix F is a fixed set of spatially dependent functions, and matrix A is a distribution of the fixed spatially dependent functions within the data matrix D.

10 87. The process according to claim 1 wherein the original data matrix D is a series of images, matrix F is a set of unvarying images and A is a measure of how the images in matrix F are distributed in data matrix D.

88. The process according to claim 87 wherein the original data matrix D is a set of images acquired at different wavelengths.

15 89. The process according to claim 87 wherein the original data matrix D is a set of images acquired at different times.

90. The process according to claim 1 wherein the data matrix D is a set of measurements representing behavioral studies.

91. The process according to claim 1 wherein the data matrix D is a set of measurements representing clinical studies.

20 92. The process according to claim 1 wherein the data matrix D is a set of measurements representing biomedical research studies.

93. The process according to claim 1 wherein the data matrix D is a set of measurements representing psychodynamic studies.